Il cloud computing e le biblioteche: illusione o opportunità?

Valdo Pasqui

Introduzione

Cloud computing è un termine diffusamente utilizzato per riferirsi a una varietà di piattaforme e soluzioni che forniscono capacità di calcolo e di memorizzazione e che implementano applicazioni condividendo un medesimo principio: le risorse e i servizi sono dislocati nella rete, spesso in modo indefinito, e forniti attraverso di essa in modo scalabile. Si tratta di una espressione che sottende una molteplicità di tecnologie ma che si riferisce soprattutto alla modalità di erogazione e fruizione dei servizi, purtroppo non sempre citata a proposito, talvolta abusata, enfatizzata negli ultimi 2-3 anni dalla maggior parte dei produttori ICT, alcuni dei quali stanno cavalcando l'onda della nuova moda anche per riproporre in veste rinnovata soluzioni preesistenti e dunque degna della massima attenzione.[1]

[1] Alla fine di settembre 2010 su alcuni quotidiani a tiratura nazionale è comparsa a piena pagina la pubblicità con la quale Telecom Italia promuove "Nuvola italiana" (http://www.nuvolaitaliana.it) la piattaforma di cloud computing con la quale si propone come fornitore di soluzioni innovative ed economicamente competitive per le imprese e la pubblica amministrazione. Il termine "cloud computing" è così arrivato al vasto pubblico di non addetti ai lavori e non esperti di ICT.

JLIS.it. Vol. 1, n. 2 (Dicembre/December 2010), p. 277–304.
DOI: 10.4403/jlis.it-4527

Anche le biblioteche, inevitabilmente sempre più legate alle tecnologie (es. i dispositivi mobile, le reti wi-fi), sono sollecitate da iniziative e soluzioni che più o meno esplicitamente si rifanno ai principi del cloud computing. Testimonianza di questa ″influenza″ è il fatto che durante la recente Conferenza annuale ALA, nel giugno 2010, la Library and Information Technology Association (LITA) ha dedicato una sessione al ″Cloud computing for library services″ e che la stessa associazione ha intitolato il 13mo Annual LITA National Forum, tenutosi ad Atlanta dal 30 settembre al 3 ottobre 2010, ″The Cloud and the Crowd″[2] dedicando una sessione principale, tenuta da Roy Tennant, a questa tematica. Si tratta di una moda passeggera? Solo una rivisitazione del ben noto modello di gestione in hosting di servizi e applicazioni? Come orientarsi per riconoscere le proposte realmente innovative dagli approcci tradizionali esposti con un linguaggio accattivante? In ambito bibliotecario quali prospettive si possono intravedere e quali rischi occorre evitare? Queste sono alcune delle domande a cui l'articolo tenta di dare una risposta richiamando i principi base del cloud computing, descrivendo alcuni esempi e cercando di individuare caratteristiche e parametri di cui tener conto per valutare soluzioni e possibili linee di sviluppo.

Cloud computing: caratteristiche e modelli

Il cloud computing ha caratteristiche e si esplicita secondo modalità non rigorosamente formalizzate e anche per questo è oggetto di un ampio dibattito tra gli addetti ai lavori, sollevando non poche diffidenze tra i manager e tra coloro che sono preposti a prendere decisioni in merito alle soluzioni applicative e alle piattaforme che si richiamano a questo approccio. Conseguentemente (pre)visioni più o meno entusiastiche come le seguenti:

[2]http://lita.org/ala/mgrps/divs/lita/litaevents/nationalforum.

> I think it is one of the foundations of the next generation of computing [. . .] Everything we think of as a computer today is really just a device that connects to the big computer that we are all collectively building;[3]

> a way to increase capacity or add capabilities on the fly without investing in new infrastructure, training new personnel, or licensing new software[4]

convivono con affermazioni completamente opposte, sovente molto negative, come queste due:

> the interesting thing about cloud computing is that we've redefined cloud computing to include everything that we already do;[5]

> the term "cloud computing" is a marketing buzzword with no clear meaning. It is used for a range of different activities whose only common characteristic is that they use the Internet for something beyond transmitting files. Thus, the term is a nexus of confusion. If you base your thinking on it, your thinking will be vague.[6]

Per cercare di chiarire alcuni concetti e fornire qualche punto di riferimento si ritiene opportuno riprendere la definizione fornita dal NIST:

---

[3]Tim O'Reilly, CEO O'Reilly Media in Defining cloud computing CNET video interview di Dan Faber, 7 maggio 2008 http://news.cnet.com/8301-13953_3-9938949-80.html?tag=mncol.

[4]Galen Gruman, Eric Knorr What cloud computing really means InfoWorld Quote, 7 aprile 2008 http://www.infoworld.com/article/08/04/07/15FE-cloud-computing-reality_2.html.

[5]Larry Ellison Oracle CEO's in Oracle's Ellison nails cloud computing di Dan Faber CNET news, 26 settembre 2008 http://news.cnet.com/8301-13953_3-10052188-80.html.

[6]Words to Avoid (or Use with Care) Because They Are Loaded or Confusing Free Software Foundation, Inc http://www.gnu.org/philosophy/words-to-avoid.html#CloudComputing.

> cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. (Mell e Grance)

Questa definizione è accompagnata da una specifica articolata secondo tre assi che esamineremo, trattandoli in ordine diverso e integrandoli con alcuni esempi e riferimenti.

Modelli di servizio

Il primo asse di riferimento comprende le tipologie e le modalità secondo le quali servizi di cloud computing sono messi a disposizione e possono essere sfruttati dagli utilizzatori ("consumer"); si distinguono tre modelli:

Infrastructure as a Service (IaaS): l'utilizzatore può usufruire di risorse quali capacità di elaborazione, memorizzazione (spazio disco) e di rete, in pratica dei server virtuali sui quali è libero di installare un proprio stack applicativo, dal sistema operativo fino alle applicazioni (i programmi). La gestione avviene in piena autonomia come se si disponesse localmente e fisicamente delle equivalenti risorse hardware e software, pur senza avere il controllo dell'infrastruttura di cloud computing sottostante, avvalendosi delle caratteristiche che essa garantisce. Amazon Elastic Compute Cloud (Amazon EC2),[7] Rackspace[8] e Terremark[9] sono tre esempi di questo tipo di servizi che consentono di istanziare computer virtuali già pronti (es. server Linux o Windows) o configurare nuove istanze utilizzabili remotamente tramite la rete secondo modalità e costi predefiniti,

[7] http://aws.amazon.com/ec2.
[8] http://www.rackspace.com.
[9] http://www.terremark.com/services/cloudcomputing.aspx.

appropriatamente documentati e formalizzati. Amazon fornisce anche servizi di memorizzazione come Simple Storage Service (S3) e Elastic Block Storage (EBS) che permettono attraverso opportune interfacce applicative (web services) di memorizzare oggetti identificati da una chiave univoca con dimensioni fino a 5GB o di creare volumi accessibili a blocchi di dimensioni da 1GB a 1TB. Ovviamente in questo modello l'utilizzatore deve possedere competenze tecnico-sistemistiche analoghe a quelle necessarie per amministrare e gestire i sistemi locali.

Platform as a Service (PaaS): saliamo a un livello intermedio, gli utenti hanno a disposizione un framework applicativo attraverso il quale possono sviluppare, testare e mettere a disposizione applicazioni, ma non hanno il controllo della piattaforma sottostante (risorse hardware, infrastruttura di rete e sistema operativo). Due esempi di questa modalità sono Google AppEngine[10] e Windows Azure Platform.[11] La prima consente di realizzare applicazioni scritte nei linguaggi Java e Phyton e di renderle operative avvalendosi anche di un servizio per la memorizzazione persistente dei dati (datastore) e di un servizio di autenticazione degli utenti (Google Accounts). La seconda è l'analoga piattaforma proposta da Microsoft, orientata allo sviluppo e all'esecuzione di applicazioni .NET da parte di sviluppatori già esperti dell'ambiente VisualStudio. Possiamo includere in questa categoria anche Amazon SimpleDB e Amazon Relational Database Service (RDS).

Software as a Service (SaaS): si passa al livello applicativo, gli utenti usano applicazioni che sono installate ed eseguite su una infrastruttura di tipo cloud computing, senza alcun controllo su di essa o sulle applicazioni tranne alcune funzioni di configurazione e si avvalgono di un browser Web o di altre tipologie di disposi-

[10] http://code.google.com/appengine.
[11] http://www.microsoft.com/windowsazure.

tivi collegati in rete. Gli esempi più tipici di questo ambito sono i servizi di tipo Web mail e quelli di supporto al lavoro cooperativo (condivisone documenti, calendario, video etc.) come Google Apps,[12] Docs[13] che gestisce la creazione e condivisione di documenti MS Office 2010 su Facebook, applicazioni di tipo Customer Relationship Management (CRM) come SalesForce[14] per la gestione dei call center, dei rapporti con i clienti e delle vendite, piattaforme di e-learning come Moodlerooms[15] che mette a disposizione il software open source Moodle sull'infrastruttura cloud della Dell. Con Google Apps Education Edition (GAEE) Google offre gratuitamente (previa sottoscrizione) alle Università e ai loro studenti una serie di servizi di comunicazione e collaborazione oltre alla posta elettronica; Microsoft Live@edu[16] è l'analoga soluzione offerta da Microsoft. Anche Oracle è recentemente entrata nell'arena SaaS con Oracle On Demand[17] che comprende Siebel CRM e varie altre applicazioni.

Caratteristiche

Nel secondo asse di riferimento sono allocate le proprietà che contraddistinguono le soluzioni di tipo cloud computing. Le prime tre si riferiscono alle modalità con le quali le risorse (di elaborazione, di memorizzazione, le applicazioni etc.) possono essere fruite (consumate) dagli utilizzatori:

- automaticamente in base alla necessità e senza intervento umano (*On-demand self-service*);

[12] http://www.google.com/apps/intl/it/business/index.html.
[13] http://docs.com.
[14] http://www.salesforce.com/it.
[15] http://www.moodlerooms.com/e-learning-strategy/cloud-hosting.
[16] http://www.microsoft.com/liveatedu/free-email-accounts.aspx?locale=en-US&country=US
[17] http://www.oracle.com/us/products/ondemand/index.html.

- in rete grazie a meccanismi standard che ne consentono l'accesso diffuso, pervasivo e indipendente dalla localizzazione attraverso l'uso di computer, iPhone, iPad etc. (*Broad network access*);

- in modo elastico garantendo la rapida scalabilità, verso l'alto o verso il basso (*Rapid elasticity*), senza porre limiti alla possibilità di acquisire dinamicamente le risorse necessarie per la gestione dei carichi di lavoro.

A queste tre proprietà di base si affiancano altre due caratteristiche che indicano come le risorse sono rese disponibili:

- in condivisione (*Resource pooling*) secondo meccanismi che consentono di supportare una molteplicità di utenti, per esempio mediante l'uso condiviso della stessa istanza di un software applicativo da parte di più utenti (*Multi-tenancy*) e l'assegnazione dinamica di risorse virtuali e fisiche sulla base del consumo effettivo, quindi un approccio che può essere paragonato ai taxi collettivi e al car-sharing nei trasporti cittadini;

- accompagnate da processi di controllo e misurazione del consumo che consentono di ottimizzare l'impiego delle risorse disponibili e che permettono di raccogliere i dati sul loro utilizzo ai fini della contabilizzazione.

L'allocazione e il rilascio delle risorse in modo elastico e gli strumenti di misurazione del loro consumo permettono ai fornitori di soluzioni/piattaforme cloud computing di implementare meccanismi di pagamento flessibili di tipo pay-as-you-go che ne costituiscono uno dei principali punti di forza. Infatti, non casualmente, il motto di Amazon WS è ''pay only for what you use''. Le caratteristiche elencate sono anche utili indicatori per valutare la reale natura di servizi che vengono proposti sul mercato come cloud computing.

La grande abilità marketing di molte aziende non facilita il compito di distinguere, specialmente nell'ambito delle applicazioni, servizi di hosting tradizionali (le applicazioni girano su una piattaforma esterna gestita da una terza parte) da quelli erogati secondo un effettivo approccio SaaS. Proprio l'elasticità, la scalabilità automatica, la disponibilità su richiesta e le modalità di pagamento a consumo permettono di riconoscere i servizi del secondo tipo e valutarne la reale vantaggiosità dal punto di vista economico.

Modalità di *deployment*

L'ultimo asse di riferimento riguarda le modalità secondo le quali i servizi cloud sono organizzati e resi disponibili (Cloud Computing Use Case Discussion Group). Il modello più diffuso, quello usato dagli esempi citati in precedenza, viene definito Public cloud poiché si riferisce a servizi messi a disposizioni degli utenti attraverso Internet e usati mediante un qualunque browser Web e altri dispositivi di rete oppure da applicazioni esterne tramite interfacce programmatiche (es. Web Services, REST etc.). I servizi proposti secondo questa modalità sono assimilabili a quelli erogati dalle cosiddette "public utilities" (energia elettrica, acqua, gas, fognature, telefonia) e infatti in letteratura vengono classificati come "utlity computing" (Ambrust 2009), un termine che rimanda suggestivamente a una definizione di John McCarthy, uno dei padri fondatori dell'informatica (inventore del LISP), che nel 1961 in occasione del MIT Centennial aveva preconizzato:

> if computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public

> utility [...] The computer utility could become the basis of a new and important industry.[18]

All'estremità opposta sono collocate le Private cloud che pur esibendo le stesse caratteristiche tecnico-funzionali sono utilizzate nel contesto di una determinata organizzazione o azienda di solito utilizzando le reti interne e garantendo maggiori livelli di sicurezza. Le Private cloud sono una delle soluzioni innovative proposte per realizzare i data center aziendali al fine di garantire scalabilità, flessibilità, alte prestazioni, sicurezza e resilienza utilizzando componenti (server, dischi, controller, interfacce di rete etc.) standard e non troppo costosi. In mezzo a queste due estremità si collocano due varianti che in gran parte derivano da esigenze di tipo organizzativo. La prima è costituita dalle Hybrid cloud che prevedono di mantenere sull'infrastruttura aziendale le applicazioni e i dati che per esigenze di sicurezza, privatezza ed efficienza sono ritenuti critici, mentre altri servizi e dati sono allocati in cloud di tipo pubblico, integrati con quelli locali mediante opportune interfacce programmatiche (API). Seguendo questo approccio molte università mantengono al loro interno la gestione dei servizi di posta elettronica per il personale e affidano a fornitori esterni, in modalità SaaS, la gestione delle caselle di posta elettronica istituzionali dei propri studenti, gestendo internamente le procedure di *provisioning* e conservando il controllo dell'autenticazione. La seconda variante è quella delle Community cloud in cui servizi di tipo cloud vengono realizzati, governati, gestiti ed erogati da un gruppo di organizzazioni appartenenti a una determinata comunità, un modello che come vedremo più avanti si adatta molto bene al contesto delle biblioteche.

[18] http://en.wikipedia.org/wiki/Utility_computing.

Un cenno alle tecnologie

Di solito i fornitori di servizi cloud computing forniscono descrizioni abbastanza generiche sulle tecnologie utilizzate e dall'esterno si ha una percezione molto parziale delle architetture interne che si riduce ancor di più salendo dai servizi IaaS alle applicazioni SaaS. Per supportare le caratteristiche illustrate in precedenza le soluzioni cloud computing si avvalgono:

- delle tecniche di virtualizzazione[19] che attraverso uno strato interposto tra il sistema operativo e l'infrastruttura fisica sottostante consentono di simulare dispositivi reali (processori, memoria, supporti di memorizzazione, interfacce di rete, firewall etc.), di creare server virtuali con varie configurazioni e differenti sistemi operativi e che all'occorrenza possono essere spostati su altri server fisici (anche in siti diversamente localizzati);
- dei principi del grid computing per le elaborazioni parallele distribuite in rete;
- dei file system distribuiti per realizzare la memorizzazione permanente dei dati assicurando resilienza e ridondanza;
- delle reti a banda larga (broadband networks) per supportare l'interazione tra i servizi e le applicazioni assicurando bassi tempi di latenza per lo scambio dei dati;
- delle tecnologie dei web services (SOAP, XML, REST, http, https etc.) per realizzare le interfacce programmatiche che consentono l'accesso e l'integrazione con i servizi esposti;

[19] Xen hypervisor è uno dei più diffusi software di virtualizzazione distribuito gratuitamente dalla comunità Xen.org come GNU General Public License. VMware, MS Hyper-V e Oracle VM sono alcuni dei sw commerciali che si contendono il mercato della virtualizzazione.

- dei principi e degli strumenti tipici delle Service Oriented Architectures (SOA) per l'amministrazione e il monitoraggio dell'intera infrastruttura cloud.

Tuttavia per poter garantire che le caratteristiche peculiari del cloud computing siano rispettate e per assicurare adeguati livelli di servizio agli utenti non basta avvalersi delle tecnologie elencate, occorrono soprattutto adeguati strumenti di configurazione, governo, monitoraggio e misurazione delle risorse allocate e utilizzate. Dunque per le organizzazioni che intendono realizzare infrastrutture "private cloud" o che vogliono avvalersi di cloud ibridi e di tipo "community" sono indispensabili software di amministrazione, gestione e controllo che espongano opportune interfacce e forniscano gli strumenti di misurazione e di contabilizzazione necessari per poter implementare le politiche di pagamento flessibili. A questa area particolarmente critica corrisponde uno dei settori del mercato ICT attualmente in forte crescita che vede la presenza di tutti i maggiori competitori mondiali (Oracle dopo l'acquisizione di SUN, IBM, HP, Dell, VMware, Cytrix etc.) e nella quale sono disponibili anche soluzioni di tipo open source quali Eucalyptus,[20] OpenNebula[21] e Apache Hadoop.[22] HP, Intel e Yahoo! hanno creato Open Cirrus[23] che promuove lo sviluppo di software open source e fornisce di un ambiente di prova aperto. Certamente questo è un ambito ancora

[20] Iniziato come progetto di ricerca del Computer Science Department della University of California, Santa Barbara, Eucalyptus ora è sviluppato da Eucalyptus systems, Inc, la società fondata da alcuni dei suoi ideatori: può essere scaricato come open source (GNU GPL vers. 3 della Free Software Foundation) da http://open.eucalyptus.com.

[21] Distribuito secondo i termini della Apache License, Version 2.0 è scaricabile da http://www.opennebula.org/about:about.

[22] Si tratta di un insieme di progetti Apache che forniscono strumenti per implementare, coordinare e gestire cluster di elaborazione, data set e file system distribuiti, http://hadoop.apache.org.

[23] https://opencirrus.org.

poco standardizzato (per esempio sia Eucalyptus che OpenNebula implementano un'interfaccia programmatica conforme a quella di Amazon EC2 che funge da standard di riferimento, ma è un prodotto commerciale) e fortemente instabile in cui occorre muoversi con estrema prudenza, tuttavia l'esistenza di prodotti open source offre l'opportunità di sperimentare soluzioni cloud per la gestione e l'erogazione dei servizi per le biblioteche in particolare da parte di organizzazioni e consorzi.

Il cloud computing e le biblioteche

Passiamo ora a esaminare alcune aree di applicazione in cui le biblioteche possono avvalersi dei servizi di cloud computing. Per comodità di esposizione si individuano alcune categorie in base alla tipologia di applicazione e ai modelli di cloud ritenuti più appropriati:

- **applicazioni SaaS di tipo generale offerte tramite cloud pubbliche**. Servizi di posta elettronica, lavoro cooperativo, calendario, gestione dei documenti, firma digitale remotizzata (ovvero senza l'uso di smart-card e dei relativi lettori) e di conservazione sostitutiva. Le biblioteche, singole, consorziate o raggruppate in reti cooperative di carattere istituzionale e territoriale, come qualunque altra impresa o Pubblica Amministrazione, possono avvalersi di questi servizi eliminando i costi di acquisizione, manutenzione e gestione delle piattaforme hardware e software necessarie e del personale con competenze informatiche;

- **risorse e servizi offerti come IaaS e PaaS**. Queste soluzioni possono soddisfare esigenze derivanti dalla necessità di sperimentare, anche temporaneamente, e sviluppare nuove

applicazioni senza investire nell'acquisizione dell'hardware necessario e senza impattare sulla gestione delle infrastrutture informatiche già disponibili. Rientrano in quest'area lo sviluppo di prototipi software nell'ambito di progetti, il test di pacchetti applicativi open source e la possibilità di disporre di risorse virtuali a costi contenuti da parte di imprese "start-up" che possono così evitare di investire nell'infrastruttura per realizzare applicazioni innovative e servizi personalizzati per le biblioteche;

- **servizi di memorizzazione permanente dei dati erogati secondo il modello IaaS sia da cloud pubblici o attraverso cloud di comunità**. In questa categoria, a un primo livello, possiamo collocare l'uso di strumenti di produttività personale come Zotero, un modulo che estende Firefox e consente di gestire collezioni personali di citazioni bibliografiche, archiviare le risorse e aggiungere annotazioni secondo i più diffusi sistemi di citazione. Aprendo un account gratuito su Zotero server è possibile riversare in rete le proprie collezioni bibliografiche per gestirle con qualunque PC o dispositivo mobile e per condividerle con altri utilizzatori. Inoltre Zotero Storage mette a disposizione degli utenti che si registrano un'infrastruttura cloud per salvare i documenti allegati alle proprie collezioni bibliografiche consentendo di accedere da un qualunque browser collegato in rete, di sincronizzare tali contenuti svincolando dall'uso di un determinato PC e di condividere le risorse archiviate nell'ambito di gruppi di lavoro. I primi 100MB sono gratuiti, poi occorre pagare secondo un tariffario pubblico (es. 5GB costano 5$ al mese). Si tratta di un servizio cooperativo molto adatto per le biblioteche universitarie ma che può trovare valida applicazione anche nel contesto delle biblioteche pubbliche. A un livello più specialistico possiamo collocare

il supporto di archivi digitali (es. istituzionali) ovvero la possibilità di avvalersi di servizi di cloud storage come Amazon S3 per supportare la memorizzazione delle risorse dei repository digitali. I moduli di memorizzazione di eprints.org[24] e di Fedora[25] prevedono una modalità di configurazione per archiviare i contenuti in S3;

- **servizi che si avvalgono di modelli IaaS e SaaS per la conservazione delle risorse digitali**. Un primo esempio è TimesMachine,[26] un servizio realizzato dal New York Times per pubblicare sul Web i numeri del giornale a partire dal primo numero del 18 Settembre 1851 fino a quello del 30 dicembre 1922 e che ha utilizzato Amazon EC2/S3 e Apache Hadoop per archiviare 405.000 immagini e 3,3 milioni di articoli (Gottfrid). Il caso più significativo è DuraCloud,[27] promosso dalla organizzazione no-profit DuraSpace che ne ha avviata la sperimentazione nel 2009 e ne prevede il rilascio operativo nel primo trimestre 2011. Si tratta di un servizio che opera secondo i principi del cloud computing interponendosi tra i repository locali (es. DSpace e Fedora) e i servizi di cloud storage di tre diversi fornitori commerciali (Amazon, Rackspace, EMC Atmos). Tuttavia DuraCloud va ben oltre l'uso di risorse di memorizzazione in modalità cloud e propone un nuovo modello per la realizzazione di servizi di digital preservation. Infatti include la replica del contenuto tra coppie di fornitori di storage, la verifica e validazione dell'integrità dei dati (fixity), servizi di visualizzazione delle immagini secondo lo standard JPEG2000, di conversione puntuale o massiva di immagini tra formati

[24] http://wiki.eprints.org/w/StorageController.
[25] https://wiki.duraspace.org/display/FCR30/Fedora+Repository+3.4+Documentation.
[26] http://timesmachine.nytimes.com/browser.
[27] http://duraspace.org/duracloud.php.

diversi e di streaming per file audio e video secondo i formati più diffusi (MP3, MP4, FLV). Uno degli aspetti significativi di questa iniziativa è il fatto che il software è disponibile come open source e può essere prelevato e sperimentato liberamente.[28] Il cloud computing offre prospettive molto interessanti per trovare un equilibrio tra i requisiti della conservazione delle risorse digitali e i costi necessari per creare le appropriate infrastrutture e gestire i processi che assicurano l'integrità, la persistenza, la disponibilità e l'accessibilità dei dati nel tempo. In considerazione della sua ben nota e ampiamente discussa complessità la digital preservation è una delle aree più adatte per lo sviluppo di soluzioni realizzate attraverso l'utilizzo di cloud pubbliche e la creazione di cloud di comunità seguendo la filosofia "utility computing";

- **applicazioni gestionali per le biblioteche**. Le caratteristiche di multi-tenancy, elasticità e scalabilità del cloud computing favoriscono l'adozione del modello SaaS per i Library Management Systems di nuova concezione (estesi alla gestione delle risorse elettroniche) consentendo di eliminare le infrastrutture informatiche locali e i costi necessari per il loro mantenimento e la loro gestione. OCLC Web Scale Library Management Services[29] è al momento l'esempio più significativo di questa nuova generazione di LMS. Dopo il lancio nello scorso mese di luglio, il sistema è attualmente in fase pilota, sottoposto ai test di alcune biblioteche[30] e OCLC dichiara che è realizzato in modo da garantire la scalabilità, l'economicità e la disponibilità, offrendo tutte le funzionalità necessarie per la gestione

[28] In base alle condizioni previste dalla Apache Licence vers. 2.0 il software è limitato all'uso locale, per la modalità hosting occorre contattare duracloud.org.

[29] http://www.oclc.org/news/releases/2010/201038.htm.

[30] OCLC ha pubblicato un'intervista con i rappresentanti di tre biblioteche sperimentatrici http://vidego.multicastmedia.com/player.php?p=o3538545.

delle attività di una biblioteca moderna, incluse l'acquisizione dei periodici elettronici, la gestione delle licenze e la definizione dei workflow dei processi. Per il momento non sono stati pubblicati da OCLC dati tecnici e informazioni approfondite per poter esprimere una valutazione tecnico-funzionale esaustiva e per verificare se si tratti di un'applicazione realmente aderente ai principi del cloud computing o solo di una tipologia evoluta di hosting, tuttavia questa iniziativa di OCLC dovrebbe indurre i principali fornitori di LMS a proporre soluzioni analoghe. Infatti ExLibris, quasi in risposta a OCLC, ha annunciato il nuovo Unified Resource Management Service (URM)[31] descrivendolo esplicitamente come un "cloud service" di tipo SaaS (Stearns), concepito per ridurre il Total Cost of Ownership (TCO).[32] Liblime fornisce la versione commerciale di Koha, uno dei più significativi sistemi di gestione bibliotecaria open source, in modalità SaaS sulla propria piattaforma cloud computing.[33] Si auspica che a seguito di queste iniziative si determini un impulso nel settore LMS/ILS che, attraverso la combinazione dell'approccio SaaS e di alcuni sistemi open source (Evergreen, Koha, Kuali OLE), faccia nascere applicazioni innovative destinate sia a migliorare la qualità dei servizi bibliotecari che a ridurne i costi di gestione;

[31] URM White Paper: http://www.exlibrisgroup.com/files/Solutions/TheExLibrisFrameworkforNextGenerationLibraryServicesOct2010.pdf.

[32] L'analisi del Total Cost of Ownership è stata introdotta alla fine degli anni Ottanta dal Gartner Group e comprende la valutazione di tutti i costi, diretti e indiretti, relativi all'acquisizione dell'hardware e del software, gli aggiornamenti, la sostituzione, la dismissione e la gestione operativa di tutti i componenti (materiali e umani) necessari all'esercizio di una soluzione informatizzata. Cfr. anche la voce di Wikipedia http://en.wikipedia.org/wiki/Total_cost_of_ownership.

[33] Liblime con Koha Express http://www.liblime.com/koha-express applica comunque una politica di pagamento a sottoscrizione annuale che quindi riconduce questa soluzione al tradizionale contesto dei servizi offerti in hosting.

- **applicazioni per la gestione del catalogo**. Il catalogo ormai va inteso nella sua accezione più vasta di insieme di informazioni relative a tutte le risorse della biblioteca, indipendentemente dalla natura, dal formato e dalla localizzazione, e come cuore dei processi di ricerca e scoperta da parte degli utenti. In questo ambito varie applicazioni SkyRiver, LibraryThing, The Open Library, Goolgle Scholar e Google Books, Summon,[34] Primo Central[35] operano "in the cloud" e pur con diverse modalità di alimentazione dei contenuti, con funzionalità diversificate e con finalità differenti, condividono un medesimo principio: si basano su degli enormi contenitori di metadati e spesso di oggetti digitali indicizzati, implementano servizi di ricerca ed espongono interfacce di rete (grafiche e web services). La creazione di grandi "pozzi di dati" (Dempsey) nei quali vengono aggregati e indicizzati dati e metadati di diversa provenienza è ormai una tendenza consolidata poiché offre la possibilità di fornire funzionalità (es. indicizzazione, facets browsing, normalizzazione, deduplicazione, aggregazione secondo il modello FRBR, riorganizzazione in base ai modelli del web semantico, aggiunta di tag, annotazioni, raccomandazioni etc.) che caratterizzano la nuova generazione dei software di ricerca. Il cloud computing con le sue proprietà di scalabilità, ridondanza e disponibilità su domanda è un approccio adatto per implementare adeguatamente questi cataloghi e per offrire servizi che ne consentano un uso efficiente e flessibile. Pozzi di metadati e risorse condivise, realizzati avvalendosi

[34]SerialsSolutions definisce Summon un "web-scale discovery service" che si basa su un indice unico ed espone un'interfaccia programmatica ai fini dell'integrazione con i sistemi locali http://www.serialssolutions.com/summon.

[35]Primo Central di ExLibris è ospitato in un "cloud computing environment", può essere integrato con i sistemi locali e le risorse aggregate e indicizzate sono rese disponibili in base alle politiche sottoscritte dalle biblioteche (Sadeh).

di software open source (es. eXtensible Catalog[36]), esposti e resi accessibili come SaaS in cloud pubbliche e di comunità e accompagnati da politiche di pagamento basate sull'utilizzo effettivo, sono soluzioni perseguibili e sperimentabili dalle biblioteche per tentare di contrastare il monopolio dei grandi editori e dei fornitori commerciali al fine di realizzare sistemi economicamente sostenibili gestiti in ambiti cooperativi.

Discussione: prospettive e criticità

La schematizzazione fornita e gli esempi citati nel paragrafo precedente mettono in evidenza che il cloud computing offre un ampio spettro di potenzialità nell'ambito dei servizi utilizzati e offerti dalle biblioteche. La biblioteca, come qualunque organizzazione che si avvale dell'ICT, necessita di infrastrutture informatiche gestite in centri di elaborazione dati (server farm o data center secondo la terminologia più aggiornata) che oltre alle attrezzature (server, dispositivi di storage, apparati di rete, impianti di condizionamento, antincendio e antintrusione etc.) richiedono anche servizi complementari (monitoraggio delle prestazioni, gestione della sicurezza, manutenzione, disaster ricovery). I costi variano in funzione della complessità e quantità dei servizi offerti e devono tener conto anche della rapida obsolescenza delle piattaforme tecnologiche e applicative, dei consumi energetici e del personale tecnico (sistemisti, specialisti di rete, amministratori di database, sviluppatori etc.) necessario per la gestione e il monitoraggio. La sempre crescente necessità di adeguamento e crescita funzionale e la continua trasformazione del contesto tecnologico impongono una costante evoluzione dei sistemi bibliotecari. L'adozione e lo sviluppo di nuove applicazioni implica spesso la necessità di allocare o acquisire nuove risorse hardware (server,

[36] http://www.extensiblecatalog.org.

dischi, interfacce e apparati di rete), di aggiornare quelle esistenti (versioni di sistema operativo e database, librerie, middleware etc.), di configurare e installare sottosistemi (istanze di database, http server e altri server applicativi, sistemi di memorizzazione, sicurezze e firewall, ambienti di sviluppo etc.). A questo si aggiunge la necessità di impegnare personale con appropriate professionalità, indispensabile per valutare i requisiti, i vincoli e le modalità di installazione delle nuove applicazioni, la cui disponibilità è frequentemente un fattore critico a causa delle attività di gestione corrente. L'evoluzione dei servizi bibliotecari informatizzati è influenzata da queste componenti e attività e non dovrebbe essere gestita secondo una logica emergenziale, come purtroppo sovente accade, bensì metabolizzata da parte del management come un fattore strutturale, costitutivo, da governare sviluppando le opportune competenze, intervenendo a livello organizzativo e utilizzando al meglio le tecnologie e le soluzioni disponibili. In questa visione la crisi economica contingente costituisce un'ulteriore spinta a valutare le possibilità offerte dal mercato ICT rivolgendo una particolare attenzione alle soluzioni che consentono di ridurre i costi infrastrutturali e gestionali dei sistemi e dei servizi. In tale prospettiva la flessibilità e il polimorfismo del cloud computing offrono vantaggi almeno su quattro fronti:

- **risparmio energetico (green IT)** – avvalersi di infrastrutture e applicazioni gestite secondo i modelli cloud computing (cloud pubbliche, ibride e di comunità) implica una riduzione dei data center che vengono concentrati presso opportune strutture comportando la diminuzione dei consumi energetici per l'alimentazione e il condizionamento. Questa opportunità ha una valenza non soltanto economica poiché deve essere apprezzata anche dal punto di vista della sostenibilità ecologica che non è limitata solo all'uso di tecnologie ecocompatibili, ma si attua pure attraverso l'ottimizzazione delle infrastrutture

informatiche impiegate;

- **riduzione dei costi** – cloud pubbliche e di comunità offrono servizi flessibili e su domanda che in ciascuna delle tre varianti (IaaS, PaaS, SaaS) comportano una consistente riduzione dei costi per l'acquisizione e la manutenzione delle piattaforme hardware e software e per il personale tecnico specialistico addetto alla gestione. In particolare le applicazioni SaaS permettono, come ogni forma di hosting, l'eliminazione di gran parte delle figure specialistiche ICT inducendo un'ulteriore diminuzione dei costi gestionali;

- **flessibilità, elasticità e utilizzo su richiesta** – Queste proprietà assicurano l'erogazione dei servizi in modo scalare riuscendo a soddisfare adeguatamente eventuali picchi di carico nel consumo di determinate risorse o applicazioni. Quando i servizi cloud sono realizzati attuando adeguate politiche di tariffazione a consumo e implementano modalità di contabilizzazione trasparenti, riscontrabili da parte degli utilizzatori (passaggio dai modelli per numero di utenti concorrenti o per canone d'uso al modello "utility computing"), possono determinare un'ulteriore riduzione del TCO, oltre a fornire vantaggi dal punto di vista prestazionale e funzionale;

- **condivisione di risorse, dati e servizi** – I principi della condivisione e della cooperazione sono uno dei cardini dei servizi delle biblioteche che storicamente hanno sviluppato standard per l'interoperabilità e promosso modelli organizzativi quali la catalogazione partecipata, il prestito interbibliotecario, il document delivery, il metacataloghi, il metasearching, i repository istituzionali e il metadata harvesting. Cloud di comunità realizzate attraverso piattaforme di cloud management open source che ospitino applicazioni open source offrono la possibilità di

implementare e erogare i servizi bibliotecari sommando ai benefici precedentemente indicati i vantaggi della cooperazione e della condivisione tra biblioteche.

Accanto a queste opportunità occorre tuttavia considerare i principali fattori di rischio del cloud computing, tra cui annoveriamo:

- **dipendenza dalla rete –** Inevitabilmente i servizi di cloud computing sono fortemente condizionati dalla disponibilità della rete in termini di banda e di assenza di interruzioni del traffico. Questo rischio è particolarmente alto nell'uso di cloud pubbliche (es. quando si usano i servizi Amazon o Google), mentre può essere mitigato nella realizzazione di cloud private e consortili attraverso un'attenta progettazione e opportuni investimenti atti a garantire un adeguato dimensionamento e sufficienti livelli di ridondanza dei collegamenti tra le varie sedi che costituiscono il cloud;

- **sicurezze –** Questo è uno dei temi che anima maggiormente il dibattito tra sostenitori e detrattori del cloud computing. I secondi stigmatizzano il fatto che l'azienda/istituzione che si avvale di servizi cloud alloca i propri dati in strutture remote condivise, si avvale di risorse elaborative (Iaas, PaaS) e applicative (SaaS) in comune con altri utenti (multi-tenant), rinuncia alla propria infrastruttura (tranne nei cloud ibridi) e quindi si espone al rischio di perdere il controllo dei propri dati e delle proprie elaborazioni.
  I primi sottolineano l'importanza che i fornitori cloud si avvalgano delle tecniche per assicurare la riservatezza dei dati e che siano sviluppati sistemi di audit per poter verificare e valutare i livelli di sicurezza dichiarati. Si potrebbe osservare che il contesto bibliotecario è meno esposto a questi problemi (fatta eccezione per i dati personali degli utenti come quelli di

circolazione, consultazione e download delle risorse) poiché la condivisione dei dati (cataloghi bibliografici, megaindici e repository digitali ad accesso aperto) è uno degli obiettivi primari perseguiti da tempo dalle biblioteche. Ai fini della sicurezza i sistemi di gestione federata dell'identità (Shibboleth,[37] OpenId[38]) e gli standard (SAML,[39] XACML[40]) sono di fondamentale importanza nell'evoluzione dei servizi bibliotecari erogati in modalità cloud;

- **cloud provider e vendor lock-in –** L'assenza di standard comporta il rischio che una volta scelto un fornitore (es. Amazon) sia poi impossibile o comunque troppo costoso cambiarlo e spostare i propri servizi sulle piattaforme offerte da altri. Per i servizi IaaS sarebbero necessari un formato comune delle macchine virtuali e la standardizzazione delle interfacce applicative (web services) esposte dai fornitori cloud. Al momento Amazon WS è uno standard de facto a cui si stanno conformando sia altri fornitori commerciali che i software open source. Nell'ambito SaaS la criticità non è limitata soltanto alla disponibilità di interfacce programmatiche che consentano di interagire con i servizi esposti dalle applicazioni cloud per integrarle con le applicazioni locali. In questo caso vi è il pericolo elevato di restare vincolati al fornitore che ospita i nostri dati e le nostre applicazioni ("vendor lock-in").
  Le biblioteche che per ridurre i costi decideranno di abbandonare le piattaforme proprietarie locali per utilizzare LMS e megaindici proposti dai produttori commerciali come SaaS sono esposte al rischio di vincolarsi ancora di più a tali fornitori

[37] http://shibboleth.internet2.edu.
[38] http://openid.net.
[39] http://saml.xml.org.
[40] http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml.

che oltre a fornire l'applicazione diventano anche i detentori e i gestori dei dati. Richard Stallman, il noto fondatore della Free Software Foundation, ha espresso al riguardo una posizione decisamente radicale attribuendo alle soluzioni SaaS le stesse conseguenze perniciose dell'uso del software proprietario:

> with SaaS, the cause is that you use a copy you don't have. [...] Don't use SaaS! Don't use someone else's server to do your own computing on data provided by you (Stallman).

Al riguardo occorre ricordare che l'open cloud Manifesto[41] enuncia principi di standardizzazione e, attraverso il gruppo di aderenti a questa iniziativa, promuove iniziative rivolte ad assicurare che i ''cloud providers must not use their market position to lock customers into their particular platforms and limit their choice of providers".
Le biblioteche dovranno porre una particolare attenzione nell'esaminare le future offerte di LMS e di servizi di aggregazione/ricerca di tipo SaaS per valutare i rischi di vincolarsi ai fornitori commerciali, un legame ulteriore che si aggiungerebbe a quelli ben noti dovuti al mercato editoriale (es. pubblicazioni elettroniche);

- **necessità di metriche** – Adeguate metriche e metodologie di valutazione sono indispensabili sia per la valutazione della convenienza a passare i propri servizi in modalità cloud sia per il corretto monitoraggio del funzionamento e del costo dopo l'adozione delle soluzioni di questo tipo.
  Attualmente le metodologie e gli strumenti sono ancora allo stato embrionale e la letteratura offre esempi di vario tipo come quelli proposti dal Dartmouth College (Goldestein), la

[41] http://www.opencloudmanifesto.org/opencloudmanifesto6.htm.

metodologia di analisi dei benefici dell'uso delle storage cloud proposta da E.Walker e W.Brisken (Walker e Brisken), la necessità di misurare la "effectiveness" del servizio, un concetto che si riferisce al livello con cui vengono soddisfatte le esigenze dei clienti, sottolineata in alcuni articoli da Klubeck («Metrics for Trying Times»;«Measuring the Cloud») e l'analisi dei modelli economici riportata nel rapporto della University of California at Berkeley (Armbrust et al.).

I servizi cloud computing implicano dunque un articolato insieme di vantaggi e di categorie di rischio che caratterizzano il rapporto fruitore/fornitore e il loro successo dipende da come e quanto i consumatori confidano nei fornitori. Lo strumento che sintetizza tale relazione è il Service Level Agreement (SLA) in cui sono definite le condizioni oggettivamente misurabili per valutare i servizi. Lo SLA contiene la descrizione di ciascun servizio e la sua definizione mediante un insieme di indicatori obiettivo che variano in base alla natura dei servizi e che misurano le prestazioni, l'affidabilità, la scalabilità, la durabilità dei dati, la sicurezza etc. Questi sono accompagnati dalla specifica di metriche appropriate per misurare se il fornitore eroga i servizi correttamente e dall'individuazione dei meccanismi di monitoraggio per effettuare l'audit. Infine lo SLA definisce le responsabilità del fornitore e dell'utilizzatore, le strategie da perseguire per individuare le soluzioni quando i termini non sono rispettati e le modalità di cambiamento nel tempo (Cloud Computing Use Case Discussion Group 2010). Nell'ambito dei servizi cloud computing SLA completi e trasparenti sono gli strumenti per tutelare i consumatori dei servizi offerti e quindi da sottoporre ad accurata valutazione prima di sottoscrivere accordi e contratti.

Conclusioni

Il cloud computing è una modalità di realizzazione ed erogazione dei servizi informatizzati che recupera e ripropone in modo innovativo concetti consolidati e ben noti in ambito ICT come l'hosting e l'outsourcing affidando alla rete (cloud) il ruolo centrale per la fruizione di tali servizi da parte degli utenti. Le tecnologie su cui si fonda garantiscono le proprietà peculiari di elasticità, disponibilità e misurabilità, ma gli aspetti organizzativi e di governo hanno un ruolo ancor più rilevante che si esplicita attraverso le diverse modalità di implementazione, gli strumenti di amministrazione, le politiche di contabilizzazione del consumo delle risorse e di utilizzo dei servizi, gli SLA e il grado di confidenza dei fruitori nei confronti dell'affidabilità dei fornitori. Le biblioteche possono trarre concreti benefici dalle soluzioni cloud computing, sia relativamente alla qualità dei servizi forniti che in termini economici come contenimento dei costi infrastrutturali e gestionali in particolare nell'ambito della digital preservation, degli LMS e degli aggregatori di risorse e metadati. I servizi cloud, particolarmente quelli di tipo SaaS, che verranno proposti dai fornitori commerciali dovranno essere valutati attentamente dalle biblioteche attraverso l'individuazione di indicatori oggettivi, prestando particolare attenzione alla reale economicità delle offerte e al rischio di vincolarsi indefinitamente a tali fornitori. Senza assumere la posizione radicalmente negativa di Richard Stallman e comunque tenendo presente l'avvertimento del CEO di Oracle Larry Ellison (per quanto ormai, a distanza di oltre due anni, Oracle sia pienamente entrata nell'arena del cloud computing) le biblioteche devono guardare alle soluzioni cloud computing con estremo interesse ma adeguata cautela soprattutto al fine di riconoscere e distinguere gli inevitabili tentativi di spacciare per modalità cloud tradizionali servizi di hosting. Non tutto quanto viene proposto "in the cloud" è riconducibile ai principi del cloud

computing e soprattutto ne garantisce le caratteristiche qualificanti. Proprio anche in considerazione di questo le biblioteche dovrebbero valutare con molta attenzione soprattutto le potenzialità e le opportunità che la combinazione del cloud computing con i software open source offre per lo sviluppo di una nuova generazione di servizi condivisi ed economicamente sostenibili da realizzare e gestire attraverso le organizzazioni consortili e gli ambiti istituzionali di riferimento.

Riferimenti bibliografici

Armbrust, Michael, et al. *Above the Clouds: A Berkeley View of Cloud Computing. Technical Report No. UCB/EECS-2009-28*. Berkeley: Electrical Engineering and Computer Sciences University of California, 2009. http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html. Online.

Cloud Computing Use Case Discussion Group. *Cloud Computing Use Cases Version 4.0: A white paper produced by the Cloud Computing Use Case Discussion Group, 2 July 2010*. http://opencloudmanifesto.org/Cloud_Computing_Use_Cases_Whitepaper-4_0.pdf. Online.

Dempsey, Lorcan. «Discovery layers - Top Tech Trends 2». *Lorcan Dempsey blog*. (2010). http://orweblog.oclc.org/archives/002116.html. Online.

Goldestein, Adam. «Calculating the Cloud: Determining the True Cost of Hosting Servers in the Cloud». *EDUCAUSE Quarterly Magazine* 33.2. (2010). http://www.educause.edu/EDUCAUSE+Quarterly/EDUCAUSEQuarterlyMagazineVolum/CalculatingtheCloudDetermining/207406. Online.

Gottfrid, Derek. «The New York Times Archives + Amazon Web Services = TimesMachine». *The New York Times*. (2008). http://open.blogs.nytimes.com/2008/05/21/the-new-york-times-archives-amazon-web-services-timesmachine/. Online.

Klubeck, Martin. «Measuring the Cloud». *EDUCAUSE Quarterly Magazine* 33.2. (2010). http://www.educause.edu/EDUCAUSE+Quarterly/EDUCAUSEQuarterlyMagazineVolum/MeasuringtheCloud/206547. Online.

—. «Metrics for Trying Times». *EDUCAUSE Quarterly Magazine* 32.2. (2009). http://www.educause.edu/EDUCAUSE+Quarterly/

EDUCAUSEQuarterlyMagazineVolum / MetricsforTryingTimes / 174591. Online.

Mell, Peter e Tim Grance. *The NIST Working Definition of Cloud Computing*. 15ª ed. Gaithersburg, MD: U.S. Government's National Institute of Standards and Technology, 2009. http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc. Online.

Sadeh, Tamar. «Primo Central in action». *The Ex Librarian Newsletter*. (2010). http://www.exlibrisgroup.com/default.asp?catid={2FFF5743-9CCF-440B-94AF-B187363EB635}&details_type=1&itemid={8568E718-F383-4FC5-8183-8FDBBF4D5552}, online.

Stallman, Richard. «Who does that server really serve?» *Boston Review*. (2010). http://www.gnu.org/philosophy/who-does-that-server-really-serve.html. Online.

Stearns, Susan. «Reducing Total Cost of Ownership with the Ex Libris Unified Resource Management Service». *The Ex Librarian Newsletter*. (2010). http://www.exlibrisgroup.com/default.asp?catid={4C58C886-0D23-4172-BA43-FDF90F410094}&details_type=1&itemid={8B095165-5597-41DB-A9AD-A2709C3E92C5}. Online.

Walker, Edward e Walter Brisken. «To lease or not to lease from storage clouds». *IEEE Computer* 43.4. (2010): 4–49. A stampa.

Informazioni

L'autore

Valdo Pasqui
Università di Firenze. Centro Servizi Informatici e Informativi dell'Ateneo Fiorentino (CSIAF)
AIB. Gruppo di studio sulle biblioteche digitali
Email: valdopasqui@tiscali.it

Il saggio

ACKNOWLEDGMENT:

Data di submission: 2010-10-10

Data di accettazione: 2010-11-09

Ultima verifica dei link: 2010-11-26

Data di pubblicazione: 2010-12-15